

密级：_____

浙江大学

硕士学位论文



论文题目 [硕士论文标题]

作者姓名 [作者]

指导教师 [导师] 教授

 [导师] 教授

学科(专业) [学科名称]

所在学院 [学院名称]

提交日期 [时间]

A Dissertation Submitted to Zhejiang
University for the Degree of
Master of Engineering



TITLE: _____

Author: _____

Supervisor: _____

Subject: _____

College: _____

Submitted Date: _____

摘要

[单击此处输入中文摘要]

关键词： [单击此处输入中文关键词]，[单击此处输入中文关键词]

Abstract

[Click here and input abstract in English]

Keywords:

[Click here and input keywords in English],[Click here and input keywords in English]

目录

摘要.....	i
Abstract.....	ii
第 1 章 绪论.....	4
1.1 研究目的和意义.....	4
1.2 研究背景和现状.....	4
1.2.1 标签语义法.....	4
1.2.2 原型理论.....	6
1.2.3 字典学习.....	6
1.2.4 正则化.....	7
1.3 研究内容.....	8
第 2 章 基于熵的模糊重构算法.....	11
2.1 基于质量函数与原型理论的模糊重构.....	11
2.2 根据模糊重构确定质量函数.....	11
2.2.1 几种参数约减方法.....	12
2.3 基于熵的模糊重构.....	15
2.4 实验.....	20
2.4.1 λ 对重构系数的影响	20
2.4.2 手写数字识别.....	21
2.4.3 纹理分类.....	21
第 3 章 [单击此处输入标题，页眉会自动更新]	错误! 未定义书签。
3.1 第一节.....	错误! 未定义书签。
3.1.1.....	错误! 未定义书签。
3.2 本章小结.....	错误! 未定义书签。
3.2.1.....	错误! 未定义书签。
第 4 章 [单击此处输入标题，页眉会自动更新]	错误! 未定义书签。
4.1 第一节.....	错误! 未定义书签。
4.1.1.....	错误! 未定义书签。
4.2 本章小结.....	错误! 未定义书签。
4.2.1.....	错误! 未定义书签。
第 5 章 [单击此处输入标题，页眉会自动更新]	错误! 未定义书签。

5.1 第一节.....	错误! 未定义书签。
5.1.1.....	错误! 未定义书签。
5.2 本章小结.....	错误! 未定义书签。
5.2.1.....	错误! 未定义书签。
第6章 [单击此处输入标题，页眉会自动更新]	错误! 未定义书签。
6.1 第一节.....	错误! 未定义书签。
6.1.1.....	错误! 未定义书签。
6.2 本章小结.....	错误! 未定义书签。
6.2.1.....	错误! 未定义书签。
第7章 [单击此处输入标题，页眉会自动更新]	错误! 未定义书签。
7.1 第一节.....	错误! 未定义书签。
7.1.1.....	错误! 未定义书签。
7.2 本章小结.....	错误! 未定义书签。
7.2.1.....	错误! 未定义书签。
参考文献.....	23
攻读硕士学位期间主要的研究成果.....	27
致谢.....	28

图目录

图 1.1 论文中图的格式要求.....错误! 未定义书签。

图 3.1 流程图.....错误! 未定义书签。

表 2.1 简单的多描述分配表.....错误! 未定义书签。

第1章 绪论

1.1 研究目的和意义

生活中人们通过自然语言交流信息。自然语言给人们提供了一个表达自己认知、思想的工具。但是自然语言中的词汇表达的语义通常具有不确定性。比如人们对年龄的认知，可以使用青年描述一个 30 岁的人，也可以使用中年描述他。我们很难从二者中选取一个标准的词汇描述。转换成数学语言就是很难用一个经典的集合理论描述此类问题。在经典集合的理论中，一个对象要么属于一个集合，要么不属于，只能描述明确的、非此即彼的问题。为解决此类模糊概念的问题，Zadeh 最先提出模糊集理论对经典集合做了有效的扩充。模糊集合中隶属度的加入使我们可以用模糊集合的方法描述此类问题。但是在解决实际问题中隶属度函数的求解却没有一个通用的计算模型。通常使用的方法为分析法、统计法、参照已知函数法。其中分析法和参照已知函数法为根据先验知识构造特定函数作为隶属度函数，此法有赖于研究人员的经验和对问题的分析，具有较强的主观性。统计法需要大量的统计实验来确定隶属度函数。

我们提出了一种基于模糊重构和原型理论的语义标签方法。

1.2 研究背景和现状

1.2.1 标签语义法

Lawry 提出的标签语义学是用以描述含糊概念的不确定性的理论。含糊概念表示人群在使用该语义标签交流时所表达的语义信息。在语义标签学中有两个基础的且具相关的两个概念，它们表示使用一个标签描述一个样本适合程度的方法。其中一个为质量函数 (mass function)，表示使用一个标签描述某一样本的不确定程度。这种不确定性与语言习惯的含蓄程度和背景知识相关。第二个方法为使用语义表达式描述一个样本的适合度 (appropriateness measures)。适合度表示一个人主观的对使用一个语义标签描述一个样本的置信度，这依赖于他的背景知识。

根据标签语义学，一个人需要描述一个样本 $x \in \Omega$ 时，他会考虑 LA (一个有限

集)中的每一个标签,然后试图从从选出一个适合描述样本的标签子集。将这个子集记为 D_x 。因为对标签具有不确定性,对由标签构成的集合 D_x 同样具有不确定性。在标签语义学中对 D_x 的不确定性有概率质量函数 $m_x: 2^{LA} \rightarrow [0,1]$ 来决定。

定义 1 (标签的质量函数)。

$\forall x \in \Omega$ 的概率质量函数 $m_x: 2^{LA} \rightarrow [0,1]$, 其中 $\sum_{S \subseteq LA} m_x(S) = 1$ 。

$m_x(S)$ 表示一个人对使用标签集 S 描述 x 的适合程度的认知,这里为方便表述将 D_x 记为 S 。

通过使用逻辑连接符将基本的语义标签连接起来,我们可以得到复杂的语义表达式表示的概念。语义表达式集合定义为:

定义 2 (语义表达式集合)

语义表达式集合 LE 通过递归的方式定义为:

$LA \subseteq LE$;

如果 $\theta \in LE, \varphi \in LE$, 则 $\theta \wedge \varphi \in LE, \theta \vee \varphi \in LE, \neg\theta \in LE$ 。

根据标签语义学断言“ x 是 θ ”是 D_x 上的直接约束。定义 $\lambda(\theta)$ 是所有的满足“ x 是 θ ”标签集。

定义 3 (λ 映射)

每一个语义表达式 $\theta \in LE$ 都是将 LA 的子集使用逻辑连接词连接起来的表达式。 $\lambda(\theta)$ 映射通过递归的方式定义为:

$\lambda(L_i) = \{S \subseteq LA | L_i \in S\}, L_i \in LA$;

$\lambda(\theta \wedge \varphi) = \lambda(\theta) \cap \lambda(\varphi)$;

$\lambda(\theta \vee \varphi) = \lambda(\theta) \cup \lambda(\varphi)$;

$\lambda(\neg\theta) = (\lambda(\theta))^c$ 。

基于 λ 映射,定义 $\mu_x(\theta)$ 为所有的 $\lambda(\theta)$ 中所有的标签的概率质量函数 m_x 的和。

定义 4 (适合度) $\forall \theta \in LE, \forall x \in \Omega$, 使用 θ 描述 x 的适合度定义为:

$$\mu_\theta(x) = \sum_{T \in \lambda(\theta)} m_x(T) \quad 1-1$$

适合度 $\mu_\theta(x)$ 和质量函数 m_x 通过 λ 映射连接起来。依据此关系可以得出适合度的计算满足排中律、不矛盾律、幂等律。对任意的 $\theta \in LE, x \in \Omega$, $\mu_{\theta \vee \neg\theta}(x) = 1$,

$\mu_{\theta \wedge \neg \theta}(x) = 0$, $\mu_{\theta \vee \theta}(x) = \mu_{\theta}(x)$ 。更多的性质请参考文献。

特别的, 对于 $\forall L_i \in LA, \forall x \in \Omega$, 对于 L_i 的计算可以写成以下形式:

$$\mu_{L_i}(x) = \sum_{L_i \in S} m_x(S) \quad 1-2$$

1.2.2 原型理论

假设标签集 LA 中每一个标签 L_i 都对应了一个高位空间中的原型 P_i 。原型 p_i 表示 L_i 所对应的典型样本。标签 L_i 的原型具有 p_i 的典型特征。

定义 5 给定一个非空的集合 $S \subseteq LA$, 语义表达式 $\alpha_S = \bigwedge_{L_i \in S} L_i$ 的原型 q_{α_S} 定义为:

$$q_{\alpha_S} = \sum_{L_i \in S} p_i / |S| \quad 1-3$$

$q_{\alpha_S} = \sum_{L_i \in S} p_i / |S|$ 定义的主要思想是 S 的原型 q_{α_S} 应该与 S 中的所有的 L_i 的原型 p_i 接近。一个非常自然的做法就是原型 q_{α_S} 为 S 中的所有的 L_i 的原型 p_i 的均值。例如, $L_i \wedge L_j$ 的原型为 $(p_i + p_j)/2$ 。

对于任意的 $L_i \neq L_j$, 语义表达式 $L_i \wedge \neg L_j$ 的原型 $q_{L_i \wedge \neg L_j}$ 应该靠近 p_i 但是远离 p_j 。所以, $q_{L_i \wedge \neg L_j} = p_i$ 。我们可以总结出任意的只包含逻辑连接符 \wedge 的语义表达式的原型。

定义 6 给定任意的非空集和 $S \subseteq LA$, 语义表达式 $\beta_S = (\bigwedge_{L_i \in S} L_i) \wedge (\bigwedge_{L_i \notin S} \neg L_i)$ 的原型 q_{β_S} 定义如公式 $q_{\beta_S} = q_{\alpha_S}$ 1-4 $q_{\beta_S} = q_{\alpha_S}$ 1-4 所示:

$$q_{\beta_S} = q_{\alpha_S} \quad 1-4$$

需要特别指出的是语义表达式 β_{\emptyset} 没有原型。使用一个原型表示语义表达式 $\beta_{\emptyset} = \bigwedge_{i=1}^n \neg L_i$ 非常困难。

1.2.3 字典学习

稀疏表示成为最近研究的热点问题, 在图像分类、人脸识别、图像去噪、压缩感知等方向应用广泛。在图像分类问题中, 通常有一组已经标注类别的训练数据集, 依据训练集将待分类图像进行正确分类。给定 k 个类别的训练样本, 第 k 类训练样本可以表示为 $A_i = \{v_{1,i}, v_{2,i}, \dots, v_{n_i,i}\} \in R^{m \times n_i}$, 公式中 $v_{1,i}$ 表示第 i 类第一个图像的向量, n_i 为第 i 类图像的个数, m 为将图像拉伸成列向量的维数。将所有类

别的数据构成一个矩阵 $A \in \mathbb{R}^{m \times n}$ 。A 中的一列 $v_{j,i}$ 为字典中的一个原子。新的训练样本 $y \in \mathbb{R}^m$ 可以表示为训练样本的线性组合：

$$\hat{w} = \arg \min_w \|y - Aw\|_2^2, s.t. \|w\|_0 < \varepsilon \quad 1-5$$

$w = \arg \min_w \|y - Aw\|_2^2, s.t. \|w\|_0 < \varepsilon$ 1-5 中 w 是 y 的稀疏表示形式。 ε 是控制稀疏性的阈值。约束条件中 l_0 范数用来限制 w 中的非零项的个数，也就是字典表示的稀疏程度， $w = \arg \min_w \|y - Aw\|_2^2, s.t. \|w\|_0 < \varepsilon$ 1-5 中，因为 l_0 范数求解问题是一个 np 难的问题，所以无法求解出最优解。然后我们使用 l_1 范数的约束替代 l_0 范数约束。然后在使用拉格朗日法，问题可以等效为公式 $\hat{w} = \arg \min_w \|y - Aw\|_2^2 + \lambda \|w\|_1$ 1-6 的问题求解：

$$\hat{w} = \arg \min_w \|y - Aw\|_2^2 + \lambda \|w\|_1 \quad 1-6$$

公式 $\hat{w} = \arg \min_w \|y - Aw\|_2^2 + \lambda \|w\|_1$ 1-6 中 λ 为正则系数，平衡重构准确度和稀疏性。在进行分类的时候可以通过判别法直接进行分类。判别函数为：

$$\text{identity}(y) = \arg \min_i r_i(y)$$

其中 $r_i(y) = \|y - Aw\|_2$ 为重构的误差。 i 为对应的类别。

1.2.4 正则化

常见的机器学习的方法中，一个核心的问题是如何解决在训练数据集上训练的模型，可以在测试集上也可以实现相同的效果。最常见的方法是通过加入正则化的方式降低测试误差。常见的正则化方法为 l_2 正则、 l_1 正则。本文使用的为熵正则。熵正则有一个有点为使参数尽可能的保证相同。

定义原始代价函数 (cost function) 为 $J(w; X, y)$ ，对代价函数增加 l_2 正则项后，求解问题的代价函数为：

$$\hat{J}(w; X, y) = J(w; X, y) + \frac{\alpha}{2} w^T w$$

使用 l_2 正则项，代价函数为：

$$\hat{J}(w; X, y) = J(w; X, y) + \alpha \|w\|_1$$

使用熵正则项，代价函数为：

$$\hat{J}(w; X, y) = J(w; X, y) + \alpha \sum_{i=1}^k w_i \ln w_i$$

公式中 w_i 一般取值为 0 到 1. 通过最小化熵的相反数, 可以是所有的 w_i 的取值尽可能的趋向于相同。保证所有的系数都参与运算。

下图中展示了不同的正则项对最优解产生的影响。为便于演示考虑二维情况 (w_1, w_2) 下的情况, 下图展示了不同的正则项的情况下 w 的最优解的情况, 细线为不加正则项时代价函数 $J(w; X, y)$ 的等高线, 粗线为正则项的等高线。 w^* 为不加正则项时的代价函数的最优解, w_0 为不同的正则项的约束下的最优解的情况。第一张图为正则项为 l_2 范数时的情况, 第二张图为正则项为 l_1 范数的情况, 第三张图为正则项为熵时的情况。正则项为 l_1 范数相较于 l_2 范数更容易与代价函数交于定点, 熵正则项则更容易使得所有的 w 趋近于相等。

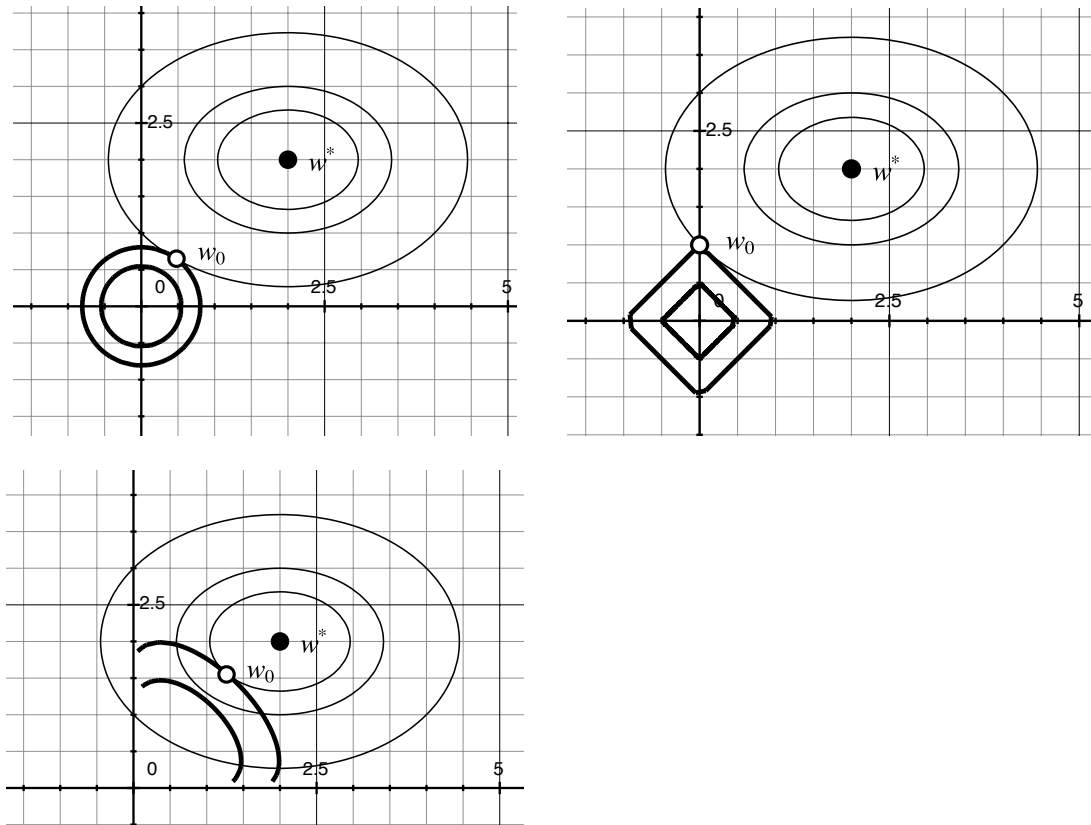


图 1-1 不同正则项情况下代价函数与正则项相交时的情况

1.3 研究内容

对于一组给定的标签 $LA = \{L_1, L_2, \dots, L_n\}$ ，每一个标签 L_i 对应了一个在空间 R^m 中的原型 $x_i \in R^m$ 。语义表达式 LE 为语义标签 L_i 通过逻辑连接符 \wedge, \vee, \neg 连接的表达式。 $L_i \in LA$ 表示空间 R^m 中“靠近 x_i ”的点，可以使用 L_i 来表示这些点。对于空间 R^m 中所有的“靠近 x_i 的 x ”都可以使用标签 L_i 描述此 x 。这是 Zadeh 在语义变量工作的一个特殊形式。

文章中使用 $\mu_{L_i}(x)$ 表示使用 L_i 描述 x 的适合度 (appropriateness)。Zadeh 提出的模糊集隶属度函数、Lawry 提出的适合度函数是两种重要的估计 $\mu_{L_i}(x)$ 的方法。但是适合度函数 $\mu_{L_i}(x)$ 必须满足每个标签 L_i 的语义适合度与标签集中的其他标签具有关联性， $\mu_{L_i}(x)$ 与其他的所有 $\mu_{L_j}(x)$ 有关。这个特性基于一个基本的认知：标签集中的标签越少，适合度函数 $\mu_{L_i}(x)$ 的值越大。例如，我们考虑一个如何从标签集中选取一个合适的标签描述一个 46 岁的人的问题。如果给定的标签集为 $LA_1 = \{\text{青年}, \text{老年}\}$ ，“老年”更适合描述此人。如果将 LA_1 扩充，加入中年，即 $LA_2 = \{\text{青年}, \text{中年}, \text{老年}\}$ ，中年更加适合描述此人。相对于 LA_1 中的适合度 $\mu_{L_i}(x)$ ， LA_2 中“青年”，“老年”的适合度函数 $\mu_{L_i}(x)$ 的值会降低。

语言作为最有效方便的交流工具，它可以描述非常丰富的信息。利用语言描述可以降低搜索的查找范围。对于一个理想的语义标签描述，我们可以使用原型 x_i 和适合度 $\mu_{L_i}(x)$ 重构出 x ， $x = \sum_{i=1}^n (\mu_{L_i}(x) x_i)$ 。

理想的语义表示应该具有以下三个性质：

1. 保序性。如果 L_i 比 L_j 更适合描述 x ，则 $\mu_{L_i}(x) \geq \mu_{L_j}(x)$ 。
2. 标签集相关性。语义适合度 $\mu_{L_i}(x)$ 的大小与语义标签集相关。
3. 重构性。 x 可以使用原型 x_i 和适合度 $\mu_{L_i}(x)$ 重构得到。

之前提出的语义标签的表示方法只具有保局部性。如 Goodman 和 Nguyen 提出的 random set 解释，Klawonn 等提出的基于相似性的框架，Lawry 和 Tang 提出的原型理论。在原型理论中，存在一个不确定的 ϵ 使得 L_i 的外延可以覆盖所有满足

$d(x, x_i) < \epsilon$ 的 x 。在最近的原型理论工作中, 将距离阈值 ϵ 表示为距离的概率密度函数 $\delta(\epsilon)$ 。使用 L_i 表示 x 的适合度 $\mu_{L_i}(x)$ 可以表示为距离的函数: $\mu_{L_i}(x) = \int_{d(x, x_i)}^{+\infty} \delta(\epsilon) d\epsilon$ 。这种表示方法只具有保局部性。

这篇文章中我们提出了一个模糊重构算法。该算法具有保序性, 标签集相关性, 重构性。模糊重构算法的提出提供了一个可以使用输入 x 和原型 x_i 利用二次规划的方法来计算适合度函数 $\mu_{L_i}(x)$ 的方法。

第2章 基于熵的模糊重构算法

2.1 基于质量函数与原型理论的模糊重构

对于任意的 $S \subseteq LA$ 都有 $m_x(S) = \mu_{\beta_x}(x)$, 所以 $m_x(S)$ 也可以度量使用语义表达式 β_S 描述 x 的适合度。因为 β_S 的原型为 q_{β_S} , 所以, $m_x(S)$ 也可以等效为使用原型 q_{β_S} 表示 x 的合适程度。也就是 $\theta \equiv \bigvee_{S \in \lambda(\theta)} \beta_S$ 。也就是任意的表达式都可以使用原子 β_S 构成。我们可以得出 x 的模糊重构。

定义 7 对于任意的 $x \in \Omega$, 它的模糊重构表示 $f(x)$ 定义为:

$$f(x) = \sum_{\emptyset \neq S \subseteq LA} m_x(S) q_{\beta_S}$$

根据模糊重构的定义, $m_x(S)$ 可以重新解释为在对 x 进行重构时, 原型 q_{β_S} 对重构的贡献程度。模糊重构 $f(x)$ 为原型 p_i 的线性组合:

$$f(x) = \sum_{i=1}^n w_i(x) p_i$$

其中 $w_i(x) = \sum_{L_i \in S} \frac{m_x(S)}{|S|}$, $i = 1, 2, \dots, n$ 。显然, $\sum_{i=1}^n w_i(x) \leq 1$, 如果 $m_x(\emptyset) = 0$ 则 $\sum_{i=1}^n w_i(x) = 1$ 。 w_i 就是 Smets 在文章中介绍的 Pignitic 概率。因为 $\mu_{L_i}(x) = \sum_{L_i \in S} m_x(S)$, 所以 $w_i(x) \leq \mu_{L_i}(x)$ 。

2.2 根据模糊重构确定质量函数

这个部分中讨论如何根据模糊重构确定任意的 x 的质量函数。实现的方法为通过最小化模糊重构表示 $f(x)$ 与原始的点 x 之间的距离来实现, 同时使用最大化熵 $-\sum_{i=1}^n w_i(x) \ln(w_i(x))$ 作为正则项, 使所有的原型都参与重构。

定义 8 给定任意的输入 $x \in \Omega$ 和基本标签 L_i 的原型 p_i , 质量函数 m_x 为如下二次规划问题的最优解。

$$\arg \min_{m_x} \left\| x - \sum_{\emptyset \neq S \subseteq LA} m_x(S) q_{\beta_S} \right\|_2^2 + \sum_{\emptyset \neq S \subseteq LA} m_x(S) \ln(m_x(S))$$

其中 $\sum_{\emptyset \neq S \subseteq LA} m_x(S) \leq 1$, 对于任意的 $S \subseteq LA$, $m_x(S) \geq 0$ 。

在上述的优化问题中，使用 l_2 范数作为重构误差进行优化。所以共有 $2^n - 1$ 个参数需要求解。为了得到质量函数的最优解，使用 m_x^j 和 q_j ，其中 $j=1,2,\dots,2^n - 1$ ，表示左右的质量函数和对应的原型。上式可以写成：

$$\arg \min_{m_x} \left\| x - \sum_{j=1}^{2^n-1} m_x^j q_j \right\|_2^2 + \sum_{j=1}^{2^n-1} m_x^j \ln(m_x^j)$$

其中 $\sum_{j=1}^{2^n-1} m_x^j \leq 1$ ，对任意 $j=1,2,\dots,2^n - 1$ ， $m_x^j > 10$ 。

2.2.1 几种参数约减方法

最小化重构误差最大化熵，可以得到质量函数 m_x 。但是问题中有 $2^n - 1$ 个参数待求解，导致求解困难，复杂度较高。下文中我们介绍几种参数约减的方法。

2.2.1.1 方法一

本小节中，我们假设对于任意的 $S \subseteq LA$ ，当 $1 < |S| < n$ 时 $m_x(S) = 0$ 。也就是语义表达式只包含 $\{\emptyset\} \cup \{LA\} \cup \{\{L_i\}: i = 1, \dots, n\}$ ，令 $m_x^i = m_x(\{L_i\})$ ，其中 $i = 1, \dots, n$ ， $m_x^* = m_x(LA)$ 。优化的目标函数在这种情况下可以重写为：

$$\arg \min_{m_x} \left\| x - \left(\sum_{i=1}^n m_x^i p_i + m_x^* \sum_{i=1}^n \frac{p_i}{n} \right) \right\|_2^2 + \sum_{i=1}^n m_x^i \ln(m_x^i) + m_x^* \ln(m_x^*)$$

其中 $\sum_{i=1}^n m_x^i + m_x^* \leq 1$ 。

这种条件的假设下，共有 $n + 1$ 的非零的质量函数需要确定。根据适合度的定义，可以根据质量函数来求解适合度 $\mu_{L_i}(x)$ ：

$$\mu_{L_i}(x) = m_x^i + m_x^*: i = 1, \dots, n$$

2.2.1.2 方法二

本小节中，我们假设对于任意的 $S \subseteq LA$ ，当 $2 < |S| < n$ 时 $m_x(S) = 0$ 。令 $m_x^{i,j} = m_x(\{L_i, L_j\})$ ，其中 $i, j = 1, \dots, n$ ，且 $i \neq j$ 。重构误差可以重写成如下形式：

$$\begin{aligned} \arg \min_{m_x} & \left\| x - \left(\sum_{i=1}^n m_x^i p_i + \sum_{i \neq j} m_x^{i,j} \frac{p_i + p_j}{2} + m_x^* \sum_{i=1}^n \frac{p_i}{n} \right) \right\|_2^2 \\ & + \sum_{i=1}^n m_x^i \ln(m_x^i) + \sum_{i \neq j} m_x^{i,j} \ln(m_x^{i,j}) + m_x^* \ln(m_x^*) \end{aligned}$$

其中 $\sum_{i=1}^n m_x^i + \sum_{i \neq j} m_x^{i,j} + m_x^* \leq 1$ 。这种假设的情况下有 $n^2 + 1$ 个参数需要求解。

2.2.1.3 方法三

假设有 LA 有一个顺序 \leq_x , 即 $L_i \leq_x L_{i-1}$ 对于任意的 $i=2,3,\dots,n$ 。令 $o_i = \sum_{j \leq i} \frac{p_j}{j}$, $i=1,2,\dots,n$ 。这种条件下只有原型 o_i 在重构 x 的过程中才使用。此时重构误差为

$$\arg \min_{m_x} \left\| x - \sum_{i=1}^n \bar{m}_x^i o_i \right\|_2^2$$

其中 $\sum_{i=1}^n \bar{m}_x^i \leq 1, \bar{m}_x^i \geq 0$ 。

这种条件下我们使用 \bar{m}_x^i 表示原型 o_i 在重构 x 的过程中的贡献程度。这种情况下我们关注所有的 m_x 为 $\{L_1, \dots, L_i\}: i = 1, 2, \dots, n\}$, $\{L_1, \dots, L_i\}$ 的质量函数为 \bar{m}_x^i 。此情况下适合度与质量函数的关系为:

$$\mu_{L_i}(x) = \sum_{j=i}^n \bar{m}_x^j: i = 1, \dots, n$$

这种情况下 m_x 为 Lawry 在他的文章中定义的质量函数, 可以通过适合度函数恢复出来:

$$\begin{aligned} \bar{m}_x^i &= \mu_{L_i}(x) - \mu_{L_{i+1}}(x): i = 1, \dots, n-1 \\ \bar{m}_x^n &= \mu_{L_n}(x) \end{aligned}$$

此假设条件下的模糊重构方法具有保序性。对于任意的 $i=2,3,\dots,n$, 都有 $L_i \leq_x L_{i-1}$ 。此方法称为保序模糊重构 (Ordering Preserving Fuzzy Reconstruction, OFPR)。

如果我们定义一个 LA 上的序关系 \leq_x 满足当且仅当 $d(x, x_i) \geq d(x, x_j)$ 时, 有 $L_i \leq_x L_j$ 。求解二次规划问题得到的模糊重构的系数的解 $[\mu_{L_1}(x), \dots, \mu_{L_n}(x)]$ 具有以下性质: 保序型, 标签集相关性, 重构性。

保序性, 上述二次规划问题中, 保序性由序关系 $L_i \leq_x L_j$ 保证。如果 $L_i \leq_x L_j$ 则 $\mu_{L_i}(x) \leq \mu_{L_j}(x)$ 。适合度 $\mu_{L_i}(x)$ 保留了 \leq_x 的序关系。有许多种适合度序关系的定义方法, 本文中的序关系 \leq_x 是基于距离定义的当且仅当 $d(x, x_i) \geq d(x, x_j)$ 时, 有

$\mu_{L_i}(x) \leq \mu_{L_j}(x)$ 。给定 x 我们可以基于标签集确定序关系 \leq_x ，将序关系 \leq_x 应用在求解 OPFR 问题中，可以得到每一个标签的适合度。序关系 \leq_x 只保留了适合度的性质，OPFR 可以得到定量的适合度同时也保留适合度的性质。

标签集相关性，适合度 $\mu_{L_i}(x)$ 与标签集内的其他标签相关。增加或者删除标签集内的其他标签都会影响基于标签集 LA 定义的适合度 $\mu_{L_i}(x)$ 的值。标签集相关性主要是因为适合度 $\mu_{L_i}(x)$ 的定义与标签集所有的内所有的原型有关。 $\mu_{L_i}(x)$ 不仅只与 $d(x, x_i)$ 有关，而且还和 $d(x, x_j)$ 有关。所以 $\mu_{L_i}(x)$ 的定义与模糊集的隶属度函数是有区别的。

语义重构性。根据适合度 $\mu_{L_i}(x)$ 和标签的原型可以重构出原始的输入 x 。语义重构性是模糊重构的最主要的特性。适合度不仅反映了输入 x 隶属于标签集中的标签的程度，而且还是一种编码方式。这个特性使得模糊重构具有很多的应用，如数据降维、模糊推理等方面。

2.2.1.4 方法四

对于任意的输入 x ，它的模糊重构输出 $f(x)$ 都可以写成以下形式：

$$f(x) = \sum_{i=1}^n w_i(x) x_i$$

其中 $w_i(x)$ 的分布满足 pignistic 概率分布。

从权重的分布 $w_i(x): i = 1, 2, \dots, n$ ，我们可以定义基于 LA 的序关系 \leq_x ，当且仅当 $w_i(x) \leq w_j(x)$ 时， $L_i \leq_x L_j$ 。通常，由 $w_n \leq \dots \leq w_2(x) \leq w_1(x)$ 时，我们定义 $L_n \leq_x \dots \leq_x L_2 \leq_x L_1$ 。我们可以得到：

$$w_i(x) = \sum_{L_i \in S} \frac{m_x(S)}{|S|} = \sum_{j=1}^n \frac{\bar{m}_x^j}{j}$$

所以 $w = A\bar{m}$ ，其中 $A = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \dots & \frac{1}{n} \\ & \frac{1}{2} & \frac{1}{3} & \dots & \frac{1}{n} \\ & & \frac{1}{3} & \dots & \frac{1}{n} \\ & & & \ddots & \vdots \\ & & & & \frac{1}{n} \end{pmatrix} \in R^{n \times n}$

$\bar{m} = A^{-1}w$, 可以得出

$$\bar{m}_x^i = i * (w_i(x) - w_{i+1}(x)): i = 1, \dots, n-1$$

$$\bar{m}_x^n = n * w_n(x)$$

同时也可以推导出 w 和 μ 之间的关系:

$$\mu_{L_i}(x) = i * w_i(x) + \sum_{j=i+1}^n w_j(x)$$

2.3 基于熵的模糊重构

基于熵的模糊重构算法

给定一个数据集 X , $X \subseteq R^{n \times d}$, 数据点之间存在较强的相似性。基于此, 可以从数据集或从数据空间 R^d 中选择一组数据原型 $P = \{p_i, i = 1, \dots, k | p_i \in R^d\}$, 使用 P 的线性组合 WP 来表示数据集 X ,

$$X = WP$$

, 其中 $W \subseteq R^{n \times k}$ 。

公式中有两个变量需要求解, 权重 W 和原型 P 。定义重构误差为 $X - WP$ 的 l_2 范式, $res = \|X - WP\|_2^2$, 其中 $0 < W < 1$, $P \subseteq R^{k \times d}$ 。熵 e 为 $-\sum_{i=1}^n \sum_{j=1}^k w_{ij} \log w_{ij}$ 。需要求解的问题为:

$$W^*, P^* = \underset{W, P}{\operatorname{argmin}} \left(\|X - WP\|_2^2 + \lambda \sum_{i=1}^n \sum_{j=1}^k w_{ij} \log w_{ij} \right),$$

$$s. t. 0 < W < 1, \sum_{j=1}^k w_{ij} = 1$$

其中第一项使得重构的数据尽可能与原始数据相似, 残差的二范式保证重构得到的数据点 WP 尽可能的与原始数据 X 在空间中具有相同的分布。第二项为熵值的罚项 e 的相反数, 最小化 e 的负数, 等价于最大化 e 。当每个权重都相同时 e 取最大值。最大化 e , 使得每个原型 p_i 都尽可能的参与重构, 而不是只有距离点 x_i 最近的几个点参与重构。如下图所示为 $k = 2$ 时, w 的取值对 e 的影响。

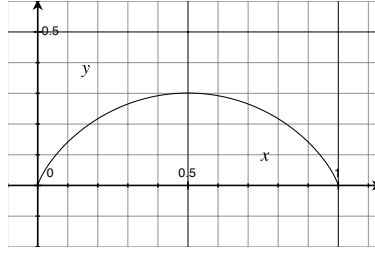


图 2-1 w 的取值对 e 的影响

对于目标函数

$$J(W) = \|X - WP\|_2^2 + \lambda \sum_{i=1}^n \sum_{j=1}^k w_{ij} \log w_{ij}, \quad 2-1$$

$$s.t. \quad 0 < W < 1, \quad \sum_{j=1}^k w_{ij} = 1$$

$$\begin{aligned} J(W) &= (X^T - W^T P^T)(X - WP) + \lambda \sum_{i=1}^n \sum_{j=1}^k w_{ij} \log w_{ij} \\ &= X^T X - X^T P W - W^T P^T X + W^T P^T P W + \lambda \sum_{i=1}^n \sum_{j=1}^k w_{ij} \log w_{ij} \\ &= X^T X - 2X^T P W + W^T P^T P W + \lambda \sum_{i=1}^n \sum_{j=1}^k w_{ij} \log w_{ij} \\ s.t. \quad &0 < W < 1, \quad \sum_{j=1}^k w_{ij} = 1 \end{aligned}$$

对上式使用拉格朗日乘数法，我们可以得到：

$$J^*(W) = J(W) + \alpha \sum_{i=1}^n \left(\sum_{j=1}^k w_{ij} - 1 \right)$$

$$\frac{\partial J^*}{\partial w_{ij}} = -(2X^T P)_{ij} + ((P^T P)^T + P^T P)_{ij}^T W + \lambda \ln w_{ij} + \lambda - \alpha = 0$$

$$\text{令 } q_{ij} = -(2X^T P)_{ij}, \quad r_{ij} = ((P^T P)^T + P^T P)_{ij}^T$$

$$\text{式可以改写成 } \frac{\partial J^*}{\partial w_{ij}} = q_{ij} + r_{ij} W + \lambda \ln w_{ij} + \lambda - \alpha$$

$$\begin{aligned}
w_{ij} &= \exp\left(\frac{\alpha - \lambda - q_{ij} - r_{ij}W}{\lambda}\right) \\
w_{ij} &= \exp\left(\frac{\alpha}{\lambda}\right) \exp\left(\frac{-\lambda - q_{ij} - r_{ij}W}{\lambda}\right) \\
\sum_{j=1}^k w_{ij} &= \exp\left(\frac{\alpha}{\lambda}\right) \sum_{j=1}^k \exp\left(\frac{-\lambda - q_{ij} - r_{ij}W}{\lambda}\right) = 1 \\
\Rightarrow \exp\left(\frac{\alpha}{\lambda}\right) &= \frac{1}{\sum_{j=1}^k \exp\left(\frac{-\lambda - q_{ij} - r_{ij}W}{\lambda}\right)} \\
\Rightarrow w_{ij} &= \frac{\exp\left(\frac{-\lambda - q_{ij} - r_{ij}W}{\lambda}\right)}{\sum_{j=1}^k \exp\left(\frac{-\lambda - q_{ij} - r_{ij}W}{\lambda}\right)} \quad 2-2
\end{aligned}$$

基于熵的模糊重构算法如下所示：

算法 1 基于熵的模糊重构算法

输入：

原型集合 $\{p_1, p_2, \dots, p_n\}$;
 需要重构的点 X ;
 设置权重衰减系数 λ ;
 误差阈值 δ ;
 最大迭代次数 \max_t 。

输出：

重构权重 W^* ;

初始化：

随机初始化权重 W^0 。

1. Repeat

2. 使用公式 $\Rightarrow w_{ij} = \frac{\exp\left(\frac{-\lambda - q_{ij} - r_{ij}W}{\lambda}\right)}{\sum_{j=1}^k \exp\left(\frac{-\lambda - q_{ij} - r_{ij}W}{\lambda}\right)}$ 2-2 更新 W^t :

$$W^t = \frac{\exp\left(\frac{2X^T P + \left((P^T P)^T + P^T P\right)^T W^{t-1}}{\lambda}\right)}{\sum_{j=1}^k \exp\left(\frac{2X^T P + \left((P^T P)^T + P^T P\right)^T W^{t-1}}{\lambda}\right)}$$

3. Until $J(W) < \delta$ or $t > \max_t$ 。

4. Set $W^* = W(t)$;

5. Return W^* 。

模糊重构算法的求解

算法:

输入: 训练数据集 X , $X \subseteq R^{n \times d}$, 原型数目 k 。

输出: 权重 W 和原型 P 。

权重学习

权重更新学习算法。固定原型 P , 求解权重 W , 目标求解问题重新定义为:

$$W^* = \underset{W}{\operatorname{argmin}} \left(\|X - WP\|_2^2 + \lambda \sum_{i=1}^n \sum_{j=1}^k w_{ij} \log w_{ij} \right)$$
$$s.t. 0 < W < 1, \sum_{j=1}^k w_{ij} = 1$$

上述函数中第一项为凸函数, 第二项也为凸函数。故目标函数为凸函数, 具有全局最优解。使用梯度下降法求解问题。

原型学习

原型更新学习算法。固定权重 W , 更新原型 P 。新的目标函数为:

$$P^* = \underset{P}{\operatorname{argmin}} (\|X - WP\|_2^2)$$

上式中的问题可以通过正规方程来求解得到结果。即:

$$P^* = (WW^T)^{-1}WX$$

上式中 WW^T 的逆可能不存在, 可以使用 WW^T 的伪逆来替代求解。循环执行 1、2 过程直到收敛, 即可求解出权重和原型。

第3章 基于模糊重构的分类

3.1 模糊 K 近邻算法 (FKNN)

K 近邻算法 (KNN) 是一种传统的分类算法, 主要通过计算与测试样本近邻的 K 个点, 然后对类别进行投票, 得到测试样本所属的类别。该算法形式简洁, 在很多方面具有应用场景。但是 K 近邻算法在类别边缘时通常无法很好的工作, 受到多个类别的干扰, 使其分类准确度下降。将模糊集理论中的隶属度引入 K 近邻算法中, 得到模糊 K 近邻算法 (FKNN)。对于任意的 $x \in \Omega$, 其隶属于类别 i 的程度记为 $\mu_i(x)$, $\mu_i(x)$ 的计算公式为:

$$\mu_i(x) = \frac{\sum_{j=1}^K \mu_{ij} (1/\|x - x_j\|^{\frac{2}{m-1}})}{\sum_{j=1}^K (1/\|x - x_j\|^{\frac{2}{m-1}})}$$

其中 m 表示距离对于对隶属度的贡献, m 增大不同距离的对隶属度区域相同; m 等于 2 时, 隶属度为距离的加权; m 变小直至等于 1 时, 距离越近的点的权重越大, 此时等效于降低 K 。其中 μ_{ij} 则是样本 x_j 相对于第 i 类的隶属度, 可以通过 $\mu_{ij} = \begin{cases} 0.51 + (n_i/K) \times 0.49 & \text{当 } L(j) = i \\ (n_i/K) \times 0.49 & \text{当 } L(j) \neq i \end{cases}$ 公式 3-1 计算出。

$$\mu_{ij} = \begin{cases} 0.51 + (n_i/K) \times 0.49 & \text{当 } L(j) = i \\ (n_i/K) \times 0.49 & \text{当 } L(j) \neq i \end{cases} \quad \text{公式 3-1}$$

其中 n_i 表示 K 个近邻中属于第 i 类的训练样本数目。 $L(j)$ 则表示样本 j 的类标签。

根据样本 x 所属类别 c , 根据隶属度 $\mu_i(x)$ 做判定, 分类抉择的判别式为:

$$c = \arg \max_i \mu_i(x)$$

第4章 实验

4.1.1 λ 对重构系数的影响

目标函数 $J(W) = \|X - WP\|_2^2 + \lambda \sum_{i=1}^n \sum_{j=1}^k w_{ij} \log w_{ij}$, 2-1 中 λ 的作用为平衡重构误差和重构系数的熵。其中重构误差项 $\|X - WP\|_2^2$ 保证重构时的精确度, 重构误差项越小表示重构精度越高, 重构系数熵保证所有的参数比较均匀, 所有的原型都尽可能的参与重构过程程度, 样本可以用所有的原型来描述。参与重构程度与重构误差之间的平衡由 λ 进行调节。每一个标签对应了一个原型图像。使用原型和重构系数可以得到原始图像的模糊重构 $f(x)$ 。如图 4-1, 为进行模糊重构的一个实例。

每个标签对应的原型



原始输入 x 重构输出 $f(x)$



图 4-1 重构过程。每一个类别对应的原型、原始输入图像、重构输出图像。图中原始重构图像为 mnist 数据集中第 6001 幅图像, 重构时 λ 取值为 0.5。

错误! 未找到引用源。中为 MNIST 数据集中重构一个手写数字在不同的 λ 的情况下 w 的取值情况。从图中我们可以清晰的看出, 随着 λ 的增大, w 的分布越来越

越均匀。图 4-2 中展示了残差随着 λ 的变化情况，随着 λ 的增大残差会越来越大，也就是重构的精度会降低。

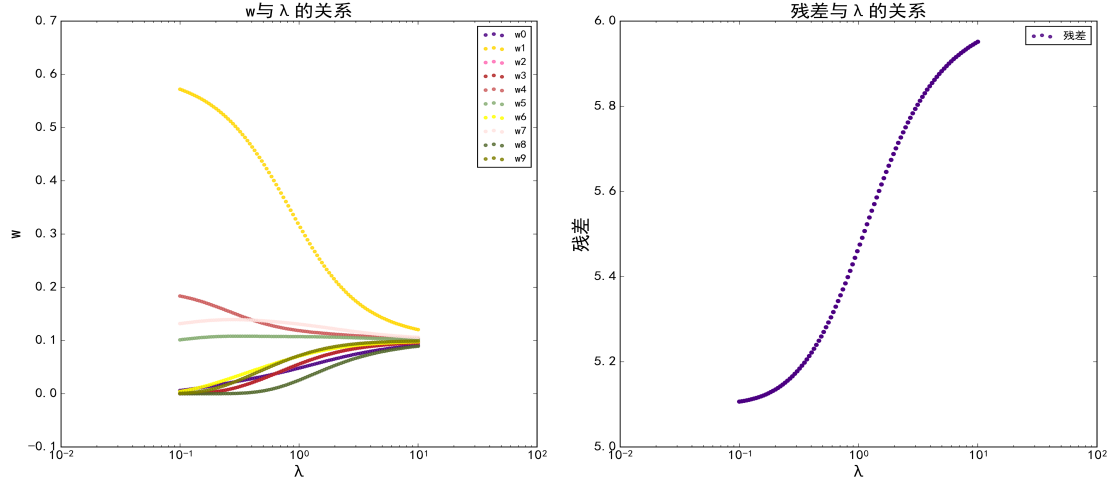


图 4-2 使用 MNIST 数据集中同一类的训练集的均值作为 P ，重构测试集中某一为 1 的手写数字图像，左图为不同的 λ 的情况下 w 的取值情况，右图为残差与 λ 的关系

4.1.2 分类实验

使用基于熵的模糊重构算法，我们可以实现分类的算法。求解出样本的原型表示后，可以使用判别式的方式进行分类的判断。样本 X_i 所属类别标签 $k = \arg \max_j W_{ij}^*$ 。

4.1.2.1 手写数字识别

本节中我们将基于熵的模糊重构算法应用在手写数字识别的问题上。我们选用了两个不同数据集，MNIST 和 USPS 数据集。MNIST 数据集包含了 70000 幅 28×28 像素的手写数字，我们使用其中的 60000 幅图像作为训练集，剩余的 10000 幅图像作为测试集。USPS 包含了 7291 幅训练图像，2007 幅测试图像。我们分别对不同的数据集进行不同的实验。实验包含不同算法间的比较。

4.1.2.2 纹理分类

我们从 Brodatz 纹理数据库中选取两类纹理图像，然后将原始纹理图像分割成 12×12 的图像块，对此图像块进行分类实验。并验证模糊重构的表达

参考文献

- [1].L. Zadeh, The concept of linguistic variable and its application to approximate reasoning, part I, *Information Sciences* 8 (3) (1975) 199–249.
- [2].L. Zadeh, The concept of linguistic variable and its application to approximate reasoning, part II, *Information Sciences* 8 (4) (1975) 301–357.
- [3].L. Zadeh, The concept of linguistic variable and its application to approximate reasoning, part III, *Information Sciences* 9 (1) (1975) 43–80.
- [4].L. Zadeh, Fuzzy sets, *Information and Control* 8 (3) (1965) 338–353.
- [5].J. Lawry, A framework for linguistic modelling, *Artificial Intelligence* 155 (2004) 1–39.
- [6].J. Lawry, *Modelling and Reasoning with Vague Concepts*, Springer, 2006. 9
- [7].D. Dubois, H. Prade, The three semantics of fuzzy sets, *Fuzzy Sets and Systems* 90 (1997) 141–150.
- [8].I. Goodman, H. Nguyen, *Uncertainty Model for Knowledge Based Systems*, North Holland, 1985.
- [9].I. R. Goodman, Fuzzy sets as equivalence classes of random sets, in: R. Yager (Ed.), *Fuzzy Set and Possibility Theory*, 1982, pp. 327–342.
- [10].H. Nguyen, On modeling of linguistic information using random sets, *Information Sciences* 34 (1984) 265–274.
- [11].F. Klawonn, R. Kruse, Equality relations as a basis for fuzzy control, *Fuzzy Sets and Systems* 54 (2) (1993) 147–156.
- [12].F. Klawonn, Fuzzy sets and vague environments, *Fuzzy Sets and Systems* 66 (1994) 207–221.
- [13].F. Klawonn, Similarity based reasoning, in: *Proc. 3rd European Congress on Fuzzy and Intelligent Technologies*, Aachen (1995) 34–38.

- [14].F. Klawonn, J. L. Castro, Similarity in fuzzy reasoning, *Mathware Soft Comput.* 3 (2) (1995) 197–228.
- [15].F. Klawonn, V. Novák, The relation between inference and interpolation in the framework of fuzzy systems, *Fuzzy Sets and Systems* 81 (3) (1996) 331–354.
- [16].J. Lawry, Y. Tang, Uncertainty modelling for vague concepts: A prototype theory approach, *Artificial Intelligence* 173 (2009) 1539–1558.
- [17].Y. Tang, J. Lawry, Linguistic modelling and information coarsening based on prototype theory and label semantics, *International Journal of Approximate Reasoning* 50 (8) (2009) 1177–1198.
- [18].Y. Tang, J. Lawry, A prototype-based rule inference system incorporating linear functions, *Fuzzy Sets and Systems* 161 (2010) 2831–2853.
- [19].J. Lawry, A methodology for computing with words, *International Journal of Approximate Reasoning* 28 (2001) 51–89.
- [20].J. Lawry, Appropriateness measures: an uncertainty model for vague concepts, *Synthese* 161 (2) (2008) 255–269.
- [21].Y. Tang, J. Zheng, Linguistic modelling based on semantic similarity relation among linguistic labels, *Fuzzy Sets and Systems* 157 (12) (2006) 1662–1673.
- [22].P. Smets, R. Kennes, The transferable belief model, *Artificial Intelligence* 66 (1994) 191–243. 10
- [23].M. Schmidt, Least squares optimization with l_1 -norm regularization, Tech. rep., CS542B Project Report (2005).
- [24].S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [25].M. Grant, S. Boyd, CVX: Matlab software for disciplined convex programming version 2.1.
- [26].J. Stoer, R. Bulirsch, R. Bartels, W. Gautschi, C. Witzgall, *Introduction to numerical analysis*, Vol. 2, Springer New York, 1993.

- [27].J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, Robust face recognition via sparse representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31 (2) (2009) 210–227.
- [28].X. Mei, H. Ling, Robust visual tracking and vehicle classification via sparse representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33 (11) (2011) 2259–2272.
- [29].A. Saeb, F. Razzazi, M. Babaie-Zadeh, Sr-nbs: A fast sparse representation based n-best class selector for robust phoneme classification, *Engineering Applications of Artificial Intelligence* 28 (2014) 155–164.
- [30].B. Zhang, F. Karray, Q. Li, L. Zhang, Sparse representation classifier for microaneurysm detection and retinal blood vessel extraction, *Information Sciences* 200 (2012) 78–90.
- [31].H. Liu, Y. Liu, F. Sun, Traffic sign recognition using group sparse coding, *Information Sciences* 266 (2014) 75–89.
- [32].Y. Tang, Z. Lee, How to adjust the distribution of nonzero elements in sparse representation: A granular locality-preserving approach, *Information Sciences* (in press).
- [33].E. Mamdani, S. Assilian, An experiment in linguistic synthesis with a fuzzy logic controller, *International Journal of Man-Machine Studies* 7 (1) (1975) 1–13.
- [34].D. Kim, C. Kim, Forecasting time series with genetic fuzzy predictor ensemble, *IEEE Transactions on Fuzzy Systems* 5 (1997) 523–535.
- [35].M. Russo, Genetic fuzzy learning, *IEEE Trans. Evol. Comput.* 4 (3) (2000) 259–273.
- [36].Y. Tang, Y. Xu, Application of fuzzy naive bayes and a real-valued genetic algorithm in identification of fuzzy model, *Information Sciences* 169 (3-4) (2005) 205–226.

- [37].R. Hyndman, M. Akram, Time series data library, <http://www-personal.buseco.monash.edu.au/hyndman/TSDL/index.htm>. 11
- [38].N. J. Randon, Fuzzy and Random Set Based Induction Algorithms, PhD Thesis, University of Bristol, 2004.

攻读硕士学位期间主要的研究成果

- [1] D.Spinellis, K.Raptis. Component mining: a process and its pattern language[J]. Information and Software Technology, 2000(42):609~617
- [2] Schinstock, D.E., Cuttino, J.F. Real time kinematic solutions of a non-contacting, three dimensional metrology frame[J]. Precision Engineering, 2000, 24(1):70-76

致谢

署名

当前日期

